

潜在変数空間でのデータ分布に関する検討

Study on data distribution in latent variable space

代 美月¹⁾ 神野 健哉¹⁾
Mizuki Dai Kenya Jin'no

概要

本研究では、画像分類タスクにおける畳み込みニューラルネットワーク (CNN) モデルの内部を変更することで精度が変化する様子を、潜在空間から確認する。具体的には、T-SNE を用いて CNN モデルの中間層の出力を比較し、各層が潜在空間に与える影響を探求します。CNN モデルは、BatchNormalization 層や Conv2D 層を重ねたり、活性化関数 Relu 層によってすべて 0 と正数に変換されることで潜在空間が変化することが予想される。この潜在空間の比較から得られる洞察を通じて、CNN モデルの学習プロセスとその精度向上メカニズムに関して検討する。

1 はじめに

近年、畳み込みニューラルネットワーク (CNN) に層を飛び越える接続であるスキップコネクションを導入した ResNet[1] や、自然言語処理で発展した自己注意機構を用いたモデルである Transformer を画像処理に応用した Vision Transformer (ViT)[2] などのモデルが画像分類のシステムとして提案されている。これらのモデルが高い分類性能を示す理由は、データ内の潜在的特徴を捉えていると推測されるが、その具体的な特定には至っていない。CNN は前半の層によってチャンネルを増やすことによって入力より次元の大きい特徴マップに変換される。その後もチャンネルを増加させていくが、畳み込み層によるストライドや Maxpooling の使用によって画像サイズが小さくなる。これによって特徴マップの次元は段々小さくなっていく。この操作によって CNN は画像の特徴を抽出していくと考えられている。特徴を抽出する時に特徴マップの次元は小さくなっていくが、分類のためには他のクラスと距離を取るために分布を大きくした方が分類精度が向上するのではないかと考える。そこで我々は、特徴マップの分布について調査するため、主成分分析を用いてデータの 70% の分散を説明するのに必要な主成分の数を求める。その値を特徴マップの次元で割ることで、各層の特徴マップに対してどのくらいの割合で分布が広がっているのかを求める。これを各層で比較することによって、分類器のどの層で分布が広がっているのかについて調査する。

2 特徴マップの調査方法

本研究における特徴マップとは、6つの畳み込み層と Global Average Pooling の出力値である。本研究では 6 層の CNN についてのみ検討を行ったため、図 1 に各畳み込み層 (c1 から c6) および Global Average Pooling (gap) によって生成される特徴マップの次元を示す。これらの層を通じて、特徴がどのように変化するかを視覚化する。これらの特徴マップをベクトル化し、主成分分析を施した。主成分分析 (Principal Component Analysis, PCA) とは、多次元データを少数の主成分に変換することによって、データの次元を削減しつつ重要な情報を保持する手法である。主成分は元の変数の線形結合として表され、分散が最大となるように選ばれる。これにより、データの特徴やパターンを視覚的に把握しやすくなる。本研究では、各畳み込み層の出力をベクトルに変換し、そのベクトルに PCA を適用して、特徴マップの 70% の分散を説明するのに必要な次元数を求める。このとき得られた次元数が、データ数よりも、特徴マップの次元よりも小さい次元だった場合、データを全て用いずに削減できることを示している。この PCA で得られた次元数を特徴マップの次元で割ることで、各層の特徴マップに対してどのくらいの割合で分布が広がっているのかを求める。これを各層で比較することによって、分類器のどの層で分布が広がっているのかについて調査する。具体的には、特徴マップの次元が小さくて、主成分分析で得られた次元数が多ければ、分布は広がっていると考えられる。

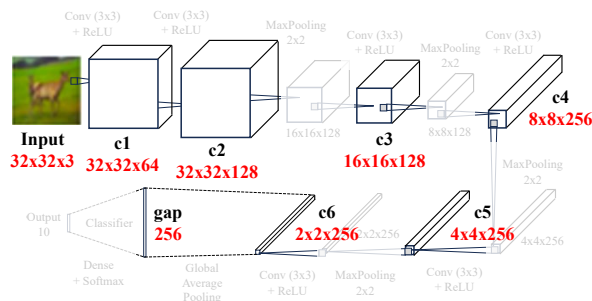


図 1 各畳み込み層 (c1 から c6) および Global Average Pooling (gap) によって生成される特徴マップ

1) 東京都市大学大学院総合理工学研究科情報専攻

3 画像分類データセット

データセットには CIFAR-10[3] を用いた。32×32 の画像サイズ、図 2 に示すように動物や乗り物で構成される 10 クラスのラベルが付与されている。トレーニング画像 50,000 枚、テスト画像 10,000 枚で構成される。



図 2 CIFAR-10

4 結果

図 3 に各モデルの特徴マップをベクトル化した時の次元数を示す。横軸が各層を示し、縦軸は次元数を示す。このグラフでは、No.1 モデルは基準モデルとした Touhu モデル [4] を表す。全てのチャンネルを 1/2 にしたモデルを No.2 とした。No.3 は c3 から c6、No.4 は c4 から c6 の特徴マップ次元が一緒になるように Maxpooling を抜いてチャンネル数を統一した。

図 4 には、PCA で得られた次元数を特徴マップの次元数で割った割合を示す。結果から、前半の層では与えられた次元数に対して実際にデータの 7 割が分布しているのは、全体の 1%以下であることがわかる。さらに、層が深くなっていくにつれて割合が大きくなっていることがわかる。特に、特徴マップの次元では c2 は c3 より大きいものに対し、割合は c3 の方が高い割合を示している。モデル No.3 から同じ特徴マップ次元である c3 から c6 において、割合が増加していることがわかる。このことから、段々と特徴マップの次元は小さくなっていても、データが分布する割合は増加していることがわかる。これは、前半で共通した特徴などを掴み、それを分類するために分布を広げているのではないかと考えられる。

さらに、この割合の値について、たくさんあるのに数%にほとんどの特徴データを格納しているとも言える。このことから、このニューラルネットワークの出力は余分な情報が多いのではないかと考えられる。この情報量について詳しく調査することで、特徴マップに広く分布するような学習方法についての検討が可能なのではないかと考えられる。

5 まとめ

本研究では特徴マップに対して大半のデータがどのくらいの次元に分布しているかを PCA によって求め、分

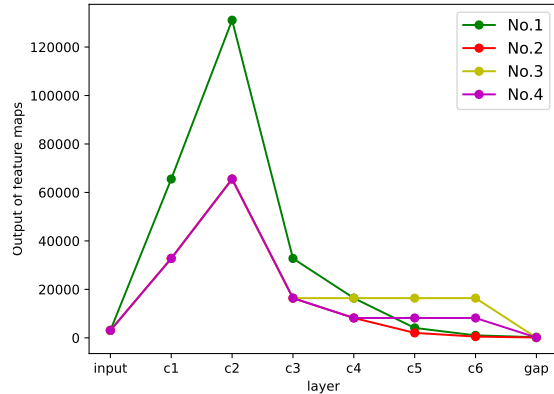


図 3 各モデルの特徴マップをベクトル化した時の次元数

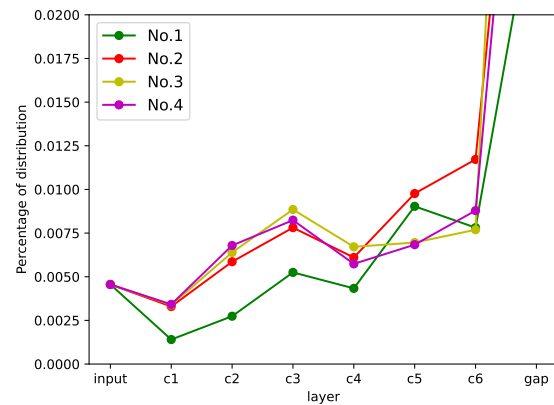


図 4 PCA の次元数とベクトル特徴マップの次元の割合

布の割合を求めた結果、層が深くなればなるほど使用する空間が広がることが示唆された。今後はモデルの層数を変化させても同じことが言えるのかについて検証を行う。

謝辞

本研究の一部は JSPS 科研費 23K11266, 23H03387, 24K15115, 東北大学電気通信研究所共同プロジェクト研究, 東京都市大学重点推進研究未来知能ユニットの助成によるものです。

参考文献

- [1] Kaiming He et al., "Deep Residual Learning for Image Recognition," Proc. CVPR 2016, pp. 770-778, 2016.
- [2] Alexey Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Proc. ICLR2021.
- [3] Alex Krizhevsky, "Learning Multiple Layers of Features from Tiny Images", 2009.
<http://www.cs.toronto.edu/~kriz/cifar.html>
- [4] Mizuki Dai, Kenya Jin'no, "Toward the realization of lightweight CNN", IEICE NOLTA2022